(1)

Feature Set Evaluation for Classifiers

FINAL REPORT

March 1, 1989

period: September 15, 1988 to March 1, 1989

contract #: DAAB07-88-C-F415

security classification: unclassified

CDRL #: B001

for: *Director*

Center for Night Vision & Electro Optics

ATTN: AMSEL - RD - NV - AC

(Martin Lahart)

Fort Belvoir, VA 22060 - 5677

by: *KAB LABORATORIES INC.*

3116 Mercer Lane

San Diego, CA 92122

(619) 455 - 5361

TABLE OF CONTENTS

	page
I. INTRODUCTION	1
A. Background	1
B. Objectives	1
C. Scope	3
II. RESULTS	
A. Chronological Summary	5
B. Detailed Results	9
1. OLPARS	9
2. Bhattacharyya Enhancement	12
3. Analysis of Results	16
III. CONCLUSIONS	22

I. INTRODUCTION

A. Background

On September 13, 1988 KAB LABORATORIES INC. (KAB) was awarded a Small Business Innovation Research (SBIR), Phase I contract with the Center for Night Vision & Electro Optics (CNVEO). The principal investigator for this research activity is John Konotchick of KAB, and the technical project manager for the work is Martin Lahart of CNVEO. Work on the contract commenced on September 15, 1988. The Phase I activity was to conduct research on feature set evaluation techniques to improve CNVEO's ability to select the best features to be used in their work of pattern recognition/classification of targets. This is the final report under that activity, and covers the entire period of the six-month contract.

B. Objectives

Automatic Target Recognizers (ATRs) have tried a wide variety of feature set classifiers in attempting to improve the quality of their classification of targets. The selection of these feature set classifiers to date has largely been based upon subjective intuition of the analyst. The analyst typically approaches the problem by starting with a proposed feature set which is derived somewhat heuristically based on an analyst's understanding of the underlying physical phenomena which differentiate a target from any background "clutter" or "noise" which may exist. This underlying phenomenology can be exceedingly complex in the case of real military targets, in real clutter filled backgrounds, imaged by electro-optical sensors under the less-than-ideal circumstances which may exist in a battle field environment.

The feature set for ATR applications could easily contain a large number of individual features or measurements (e.g., location of hot spots, geometric ratios, areas, perimeters, texture mixture, etc.). For real time systems, these features must be extracted quickly and processed to determine the target identification (classification). To minimize computations and keep ATR processor requirements at a reasonable level, the ATR algorithms should be efficient and extract only those features which are most useful to the identification process. The selection of this set of reduced features which possess the most powerful discriminating capability is the subject of this study.

KAB had proposed to use an existing software package, developed by PAR Government Systems Corporation (PGSC), called the On-Line Pattern Analysis and Recognition System (OLPARS) as a tool for feature set analysis. By using the OLPARS in our research we would be taking advantage of considerable previous work on this subject. The OLPARS was initially developed in the early 1970's as a pattern analysis support tool. Since that time it has been enhanced to increase its capability for analysis and display and to make it user friendly. It also comes with full supporting documentation. Under this contract CNVEO was to be furnished with an OLPARS licence, software, and documentation. The OLPARS was also to be enhanced by our research to include a new promising feature set evaluation algorithm aimed at meeting specific CNVEO needs.

The Phase I SBIR activity proposed meeting the following five technical objectives:

1. identify and propose a collection of feature set evaluation algorithmic tools which address unique characteristics of feature sets used in ATR applications.
2. implement at least one new promising feature set evaluation algorithm in FORTRAN and integrate it within the On-Line Pattern Analysis and Recognition System (OLPARS),

which is an existing commercial software system which provides general purpose feature set evaluation and classifier design capabilities.

3. demonstrate the performance of the new feature set evaluation algorithms already within OLPARS using feature sets derived from both real and simulated E/O imagery.

4. provide DoD with a licenced VAX-compatible copy of the augmented OLPARS software package.

5. document the proposed new feature set evaluation algorithm and the test results obtained with the newly implemented algorithm within a final technical report.

These objectives have been met, as Section II will describe. Upon completion of the Phase I activities CNVEO now possesses an independent capability to analyze, select and test feature sets and to evaluate their relative discriminating power for target classification. This capability provides a means for both improving and testing their own ATR approaches and for evaluating the approaches suggested by industry. The added enhancement also provides a capability to calculate error bounds on classification capability. The major goal of the objectives in Phase I was to determine whether feature set evaluation aids could be provided to CNVEO to enhance their ability to select features for pattern recognition/classification. This report will describe the effort and results in meeting that goal.

C. Scope

This report covers the six-month, Phase I SBIR study. The Phase I activity included $25,000 of material cost for the purchase of OLPARS, computer time, and a subcontract to PGSC for 75 man-hours of support on the OLPARS program. The remaining $25,000 was spread over 6 months for KAB manpower to support research on a CNVEO specific enhancement to OLPARS, and for incidental costs such as travel.

In the sections which follow the Phase I results and conclusions will be documented. Section II will first present a chronological discussion of significant events during the six-month effort, and will then present the detailed results of the research. Finally, Section III will present conclusions and recommendations resulting from that research.

II. RESULTS

A. Chronological Summary

When KAB LABORATORIES INC. was notified in September 1988, that its Phase I proposal had been approved, they called the Center for Night Vision and Electro Optics (CNVEO) program manager for this effort, Mr. Martin Lahart, to obtain further direction for the research. By coincidence, he was to be visiting our area in the near future. Mr. Lahart, was visiting San Diego for another purpose in late September. We took advantage of this opportunity to give Mr. Lahart a brief tutorial on OLPARS and a demonstration of the system. We also obtained further detail on CNVEO's primary areas of interest. Armed with this information we obtained and reviewed a number of research papers pertaining to their work. This research, carried out on reports from Mr. Lahart, from the Naval Ocean System Center library, and from the University of California San Diego libraries enabled us to focus on the primary needs of the CNVEO.

A second meeting with Mr. Lahart was held on October 27, 1988 at the CNVEO, Fort Belvoir, VA. The principal investigator, John Konotchick, and a PGSC representative, David Robbins were in attendance. At CNVEO request, Mr. Robbins presented an overview briefing of the OLPARS to a number of Center personnel. Following the briefing, Mr. Lahart provided us with a description of CNVEO equipment we might interface with, and also a list of the key areas of OLPARS enhancement of most interest to CNVEO. Our purpose in the visit was to be responsive to the desires of CNVEO and so this list, rather than our own would be used to select a feature set evaluation algorithm for development. The list included six possible enhancements, as follows:

1. Computation of error rates using assumed distribution and error rates;

2. Geometric transformation - How are features and error rates changed?;

3. Identifier for particular points in feature space - Mechanize an interface with ORACLE;

4. Provide a four-dimensional display of a form discussed at CNVEO;

5. Analyze relative discrimination ability of pairs of features;

6. Provide a metafile for plotting - to generate hard copy and displays.

These six possible enhancements had been discussed either in the OLPARS briefing meeting, or privately with Mr. Lahart, and were commonly understood by the KAB Team and Mr. Lahart. We were to study these and report back on which, if any, could be implemented during Phase I.

After considerable discussion and analysis by PGSC and KAB it was decided to attempt to implement #5. on the list. Algorithms analyzing pairs of discriminators had never been tried on OLPARS, but it was felt that it would add a powerful addition to the planned CNVEO capability.

The OLPARS system provides a number of discriminants for ranking an individual feature's ability to discriminate a class from all others, or ranking a feature's ability to discriminate between two classes. It does not, however, have the ability to rank "pairs" of features for their ability to discriminate classes. The enhancement which was attempted under the Phase I research effort was to provide this capability to the CNVEO system. If successful it would provide not only the ability to choose best "pairs" of features, but best combinations of features, and to provide error bounds on their classification ability.

A critical measure of the ability for feature pairs to discriminate classes is their probability of misclassification. The exact calculation of this error is often impractical or impossible, however, and so other related measures are often chosen. The most common approach is to define a separability measure, or distance, between the probability distributions of the classes under investigation. Assuming that the most important characteristic of this distance measure is its upper bound on error (of misclassification), we can rank feature pairs by their ability to minimize this error. This implies a distance measure with a known relationship to an error upper bound. A number of distance measures for these feature pairs have been derived (e.g., Matusita's, Vajda's entropy, Devijver's Bayesian distance, Ito's measure, Komogorov's variational, Toussaint's, etc.), but the Bhattacharyya distance is one that both provides a reliable measure, and one which could be easily implemented on the OLPARS.

The Bhattacharyya distance will provide a measure of which pairs of features have the highest separability between classes. All possible feature pairs can then be examined to determine their relative ability to discriminate between all possible class pairs. The Bhattacharyya distance measure will also permit any number of features to be evaluated for their ability to separate class pairs. This, as will be shown in the analysis, provides a very powerful feature set evaluation tool.

The original Phase I schedule called for the enhanced OLPARS to be delivered to CNVEO at the end of Phase I. After our visit on October 27th, however, we were asked if the basic OLPARS could be provided as soon as possible to CNVEO. KAB discussed this with PGSC, and received their approval to install OLPARS in the week of November 14-18, 1988. The quick reaction response of PGSC is the more laudatory because they scheduled the installation before either preparing the licence agreement for CNVEO or the invoice for KAB. This early OLPARS delivery, while causing minor schedule and plans changes, did not affect major schedule milestones.

A large number of research papers and reports on the Bhattacharyya distance measure were reviewed by KAB during October and November of 1988. This material was used to characterize the properties of the Bhattacharyya enhancement to OLPARS and to provide equations for the implementation of the enhancement on OLPARS. The programming of the enhancement had been scheduled for December, but some difficulties encountered delayed this implementation slightly. The OLPARS, while a mature and capable analysis system does not permit easy modification of its software. The system, moreover, is protected by licencing agreements so that configuration management of the software is important. KAB's subcontractor, PGSC, was required under the subcontract to program the Bhattacharyya distance algorithm into their OLPARS. The limited number of individuals with this skill in PGSC, became a problem. Mike Koligman is the PGSC expert on OLPARS in San Diego, but his demand on other PGSC commitments in November and December made him unavailable for support of this program. Once those commitments were behind us rapid progress was made in January.

KAB developed a simple data set, and programed a Bhattacharyya implementation using its Lotus 123 for a check on the OLPARS implementation during January. This was used during the latter part of January and early February to debug the enhancement, and to give a measure of confidence in its results. Following the checkout with the simple data set, an actual feature set on the OLPARS, the NASADATA set, was used for a detailed comparison against other OLPARS feature set evaluation techniques. The results of this evaluation are presented in the next section.

Following a successful installation and evaluation of the Bhattacharyya distance measure enhancement on the PGSC OLPARS in San Diego it was now ready to be transferred to the CNVEO OLPARS. The enhancement code, operating instructions, and final report will be delivered to CNVEO, Fort Belvoir at the final briefing/meeting on Phase I in March 1989.

B. Detailed Results

1. OLPARS

As mentioned in the previous section, the PGSC On-Line Pattern Recognition System (OLPARS) was delivered to CNVEO in November of 1988. It was followed up with telephone contact and visits by PGSC personnel in following months to insure that it could be used by CNVEO personnel. Since its delivery, Center personnel have been using the OLPARS.

OLPARS is a commercial software package which PGSC licenses for a fee. It is coded in FORTRAN 77 and runs on VAX computers under the VMS or Micro VMS operating systems. OLPARS is compatible with TEKTRONIX 4100-series, DEC GPX Graphics Workstation, and RAMTEK 9400-series color graphics displays. This powerful statistical pattern recognition and classification software system provides a flexible user interface and menu-driven command set.

The three major components of the OLPARS package are as follows: Data Structure Analysis, Measurement Evaluation, and Decision Logic. The Data Structure Analysis portion provides a variety of aids to assist the analyst in understanding the data being studied. It includes a variety of powerful graphics programs, allowing the data or subsets of the data to be viewed in two or three space color displays. These displays include:

1. <u>Coordinate Projection</u>- This projects the data onto two user-selected axes.
2. <u>Eigenvector Projection</u>- This projects the data onto the plane defined by the two largest eigenvectors computed from the covariance matrix formed by the entire data set or subset being examined. These eigenvectors show the directions of maximum variance in the data.
3. <u>Optimal Discriminant Plane</u>- This projects the data onto the plane which jointly maximizes between-class distance and minimizes within-class scatter.
4. <u>Non-Linear Mapping</u>- This maps from L-space to 2-space in such a manner as to preserve feature vector to feature vector distances.
5. <u>2-D Histogram</u>- This presents a three dimensional display of x,y verses a z which displays the number of x,y vectors in a bin of user-selected size.
6. <u>Waveform Analysis</u>- This enables feature vectors to be displayed as waveforms.

The graphics aids are used in concert with other elements of OLPARS to gain greater insight into the data.

Another component of OLPARS, Measurement Evaluation, contains a variety of analysis aids for processing the data sets and evaluating relative feature strengths. The Bhattacharyya distance measure discussed in the next section will be one of those techniques in the future. The major techniques currently in OLPARS include the following:

1. Discriminant Measure- This has three simple measures of numerical figure-of-merit for the ability of an individual feature to separate one class from all others, separate all classes, and separate specific class pairs.

2. Fisher Pairwise Discriminant- This technique is based on computing optimal linear discriminants on a class pair basis for all possible class pairs. It will order the features, taken individually, that best discriminate individual classes or class pairs.

3. Probability of Confusion- This technique, which may be used when unimodal assumptions are not justified, computes figures-of-merit similar to the Discriminant Measures using more sophisticated probability density estimation techniques.

The third major component of OLPARS is the Decision Logic portion. This portion provides mathematical and interactive graphic techniques to enable the analyst to tailor the decision logic or classifier design to fit the actual structure of the class data. Logic design is distribution free in that the design technique does not require knowledge of data class distribution type nor of the statistical independence of the features. OLPARS has the following logic types available within its program:

1. Nearest Mean Vector- A given feature vector is placed in the class for which the distance from the vector to the class mean is smallest.

2. Mahalanobis Distance- A given feature vector is placed in the class for which the class covariance-weighted distance from the vector to the class mean is smallest.

3. Fisher Pairwise Logic- A given feature is associated with a particular class based on the results

of computing optimal linear discriminants and thresholds to distinguish between every pair of classes. The pairwise decisions are combined to produce a final decision.

4. Eigenvector Method- The analyst interactively sets-up classification regions on an eigenvector projection.
5. User Modifications- The analyst can customize the above logic types by incorporating piecewise linear decision boundaries and by establishing reject regions.

The above description of OLPARS is a summary, top-level presentation of its capabilities. The complete OLPARS documentation should be reviewed for a more comprehensive description of its capabilities. It does not, however, have the capability of ranking more than one feature taken at a time. The Bhattacharyya distance measure, which was investigated under this program, does have that capability. It can choose the best combination of features, taken in any grouping, and form a ranking. The OLPARS enhanced with this Measurement Evaluation aid will permit the analyst to find the best "n" out of "L" features, and also to bound the classification error when choosing between classes. As a result of this Phase I activity, CNVEO will have an enhanced OLPARS with extremely powerful analysis capability.

2. Bhattacharyya Enhancement

The feature set evaluation algorithm chosen for implementation was the Bhattacharyya distance measure. The Bhattacharyya coefficient is defined as $b = \int [p(x:W_1)p(x:W_2)]^{1/2}dx$, and the Bhattacharyya distance as[1][2]

$$B = -\ln b = -\ln \int [p(x:W_1)p(x:W_2)]^{1/2}dx,$$

where $p(\mathbf{x}:W_i)$ is the multivariate probability density function when pattern vector $\mathbf{x}$ $(x_1,x_2,\ldots,x_n)$ belongs to class W_i $(i=1,2)$. If our class density functions are assumed to be Gaussian distributed, i.e.,

$$p(\mathbf{x}:W_i) = [1/[(2\pi)\{\det C_i\}^{1/2}]]\exp -1/2[(\mathbf{x}-m_i)^T\{C_i\}^{-1}(\mathbf{x}-m_i)],$$

where m_i is the mean of class i and $\{C_i\}$ is the covariance matrix of class i, then the Bhattacharyya distance between class E and class F will be given by,[1][3][4][5]

$$B = 1/8(m_E-m_F)^T\{(C_E + C_F)/2\}^{-1}(m_E-m_F) +$$

$$(1/2)\ln[\det\{(C_E + C_F)/2\}/[\det\{C_E\}^{1/2}\det\{C_F\}^{1/2}]],$$

where $\det\{C_E\}$ is the determinant of the covariance matrix of class E. This equation for B was implemented in the OLPARS under this program. The expression for the Bhattacharyya distance can be used to obtain a ranking of various combinations of features, (i.e., where 1, 2, ..., n features are used) for their ability to discriminate between any two classes E and F. The larger the B distance, the better will be our discrimination. It is also possible to use the Bhattacharyya distance measure to obtain a measure of the error expected from our feature selection.

The conditional Bayes error probability for a two class problem is given by,[3]

$$e*(\mathbf{x}) = \min[P(W_1:\mathbf{x}),P(W_2:\mathbf{x})],$$

where $\mathbf{x}$ = unknown pattern vector, W_i = class (1 or 2), and $P(W_i:\mathbf{x})$ = the a posteriori probability of $\mathbf{x}$ belonging to class W_i.

Using a geometric mean inequality

$$e^*(\mathbf{x}) \leq [P(W_1:\mathbf{x})P(W_2:\mathbf{x})]^{1/2}.$$

Taking the expectation of this yields,

$$E^* = e^*(\mathbf{x})p(\mathbf{x})dx \leq \int \{P(W_1:\mathbf{x})P(W_2:\mathbf{x})\}^{1/2}p(\mathbf{x})dx$$

$$\leq [P_1P_2]^{1/2} \int \{p(\mathbf{x}:W_1)p(\mathbf{x}:W_2)\}^{1/2}dx = [P_1P_2]^{1/2}b,$$

where P_1 is the a priori probability of class 1, $p(\mathbf{x}:W_1)$ is the multivariate probability density function (Gaussian in our case) of pattern vector $\mathbf{x}$ given class 1, and
<u>b = the Bhattacharyya coefficient</u> $= \int \{p(\mathbf{x}:W_1)p(\mathbf{x}:W_2)\}^{1/2}dx$. If a priori probabilities are not known, as they often aren't, a less tight bound of 1/2 can replace $[P_1P_2]^{1/2}$.

The expectation can also be written as,

$$E^* \leq \{P_1P_2\}^{1/2} \exp(-B) \leq (1/2)\exp(-B) \quad \text{where}$$

<u>B = Bhattacharyya distance</u> $= -\ln b$.

This gives the upper bound on error. Similar reasoning can derive a lower error bound for the Bhattacharyya distance measure of,

$$(1/2)[1-\{1-4P_1P_2\exp(-2B)\}^{1/2}],$$

so that we can bracket an upper and lower bound on expected error of,[3][4]

$$(1/2)[1-\{1-4P_1P_2\exp(-2B)\}^{1/2}] \leq E^* \leq [P_1P_2]\exp(-B) \leq (1/2)\exp(-B).$$

This simple error bounding provides one of the advantages of the Bhattacharyya distance measure. Through a simple analytical computation the bounds on average Bayes risk can be determined (or alternatively, 1-E*, the probability of correct classification).

The Bhattacharyya implementation developed for the OLPARS is technically only valid for Gaussian distributed classes. It can be worked out for other distributions, and in general it has been worked out for exponential density distributions, (e.g., Poisson, Gaussian, etc.)[1]. The difficulty of reprogramming OLPARS, however, does not make this a good testbed to experiment directly with various algorithms. Our assumption of Gaussian distributions is probably a fair one, however, given the current knowledge of the features to be investigated. The use of this implementation on non-Gaussian data sets moreover will still, in general, provide useful relative measures of feature classification strength. The absolute error measures, however, will not be accurate under those conditions. The strength of the Bhattacharyya enhancement to OLPARS was evident in the analysis performed on the NASADATA set, which is not a pure Gaussian data set.

references

[1] Kailath, T., "The Divergence and Bhattacharyya Distance Measures in Signal Selection", IEEE Transactions on Communications Technology, COM-15, 52, February 1967, pp 52-60

[2] Andrews, H.C., Mathematical Techniques in Pattern Recognition, John Wiley & Sons, New York, 1972, pp 37-43

[3] Devijver, P.A. and Kittler, J., Pattern Recognition: A Statistical Approach, Prentice-Hall Intl., London, 1982, pp 57-58, 97-99

[4] Chen, Chi-hau, Statistical Pattern Recognition, Spartan Books, Rochelle Park, New Jersey, 1973, pp 57-60, 71

[5] Fu, K.S., Applications of Pattern Recognition, CRC Press, Boca Raton, FL, 1981, pp 88

3. Analysis of Results

To provide a check on OLPARS during the programming of the Bhattacharyya distance measure, KAB developed a simple experimental data set which could be programmed on their own equipment. Table 1. presents the data used for this purpose. It describes a 3-class, 3-vector problem, which can be easily visualized, and calculated. A Bhattacharyya calculation was programmed on LOTUS 1-2-3, using this data, and means, covariance, inverse covariance, B, error statistics, etc. were printed out to use in debugging the OLPARS implementation. The first OLPARS implementation did have some bugs, but using this check they were quickly uncovered and corrected. The independent check thus both served to help in the implementation of the Bhattacharyya enhancement, and in our confidence in its results.

CLASS #1	CLASS #2	CLASS #3
30,49,51	53,31,72	58,60,21
40,60,56	62,40,70	65,61,50
47,49,47	55,48,72	73,66,30
50,42,53	59,56,68	80,64,10
54,58,50	61,70,60	83,70,45
	67,50,71	

TABLE 1. EXPERIMENTAL DATA SET

Plotting the data in Table 1. gives an impression that feature 3 will be a good discriminant. This indeed is the case. Using OLPARS standard analysis features we find that the Fisher Pairwise Discriminant (F) measure, and the Discriminant Measure (D) yield the results of Table 2.

DISCRIMINANT MEASURE					FISHER PAIRWISE DISCRIMINANT				
RANK	M#	VALUE	CL.	PAIR	RANK	M#	VALUE	CL.	PAIR
1	3	8.7436	B	AB	1	3	4.4015	B	AB
2	1	4.2843	A	AC	2	1	2.5533	C	AC
3	2	1.9859	C	AC	3	2	2.1806	C	AC

TABLE 2. OLPARS RANKING OF FEATURES

Although, the Bhattacharyya distance measure would not normally be used to rank individual features, it can be used here as well. We chose to select one feature out of the three, and ran the Bhattacharyya program. The Bhattacharyya distance measure ranked the features in order as 3, 1, 2 also. Table 3a. presents the data for OLPARS, and for the KAB check calculations for only one of three conditions that OLPARS would normally print (i.e., only feature 1 is presented in the table, although the data on feature 2 and 3 are also available). Table 3 illustrates the agreement between the independent calculations. Table 3b. presents similar data for the situation of using all three feature vectors. It should be noted that because of the large number of Bhattacharyya numbers that would be involved in most problems, and because this is meant to be a feature selection tool, OLPARS nominally prints out only the sum of the calculations of "exp(-B)" which is a measure of error upper bound. This is a number which will indicate which feature (or combination of features) is best for separating all classes considered. Clearly, by modifying the initial conditions for OLPARS to take only two classes at a time we could get all the values if we wanted to take the time.

CONDITION	CLASSES	KAB e^{-B}	OLPARS e^{-B}
TFF	AB	0.48747	
	AC	0.29784	
	BC	0.62766	
	SUM ALL	1.41297	1.41297

TABLE 3a. USING ONLY FEATURE # 1

CONDITION	CLASSES	KAB e^{-B}	OLPARS e^{-B}
TTT	AB	0.00416	
	AC	0.05121	
	BC	0.03762	
	SUM ALL	0.09299	0.09299

TABLE 3b. USING ALL THREE FEATURES

A summary of the best features to separate all classes for conditions of one, two and three features taken at once is presented in Table 4. It should be noted that the values related to classification error are steadily decreasing as we use more feature information. It can also be observed that for the condition of one feature used at a time, the ranking of features is the same as we had observed (i.e., 3,1,2) with the OLPARS measures in Table 2. For the condition of pairs of features taken together, features 2 and 3 are best, closely followed by features 1 and 3, and features 1 and 2 are much worse. It is interesting to note that using the feature ranking of Table 2, we might have expected the features 1 and 3 to be the best pair.

CONDITION	CLASSES	OLPARS e^{-B}
TFF	ALL	1.41297
FTF	ALL	1.88071
FFT	ALL	0.61758
TTF	ALL	0.85582
TFT	ALL	0.24246
FTT	ALL	0.23389
TTT	ALL	0.09299

TABLE 4. RESULTS SAMPLE DATA SET

Having the independent calculation confirmation of the KAB numbers for this experimental data set we now had the confidence in our implementation on OLPARS. The initial analysis, moreover, had produced results which showed encouraging potential for the Bhattacharyya enhancement. We were now ready to analyze a realistic set of data. It was decided that the NASADATA set on the OLPARS would be excellent for this purpose, because it was also on the CNVEO system, and because its characteristics had been analyzed extensively. It's only disadvantage was that it wasn't known whether it was pure Gaussian. This would mean that the absolute values of the error bounds could not be relied upon. We didn't expect this to greatly affect the performance of the Bhattacharyya implementation, however, because we are relying primarily on relative numbers for our ranking. The OLPARS implementation of the Bhattacharyya distance measure sums the results of analyses of all pairs of classes, to find the best "n" out of "L" features that will separate all classes. As will be shown, on the NASADATA set it significantly outperformed the feature sets that other OLPARS measures would have initially led us to try.

The NASADATA set has 7 classes and 12 features. Using this data set, we first looked at the best 4 features taken one at a time, as the OLPARS Discriminant Measures and Fisher Discriminant Measures do. The top 4 features given under this procedure were as follows:

Fisher 6, 10, 1, 2
DSCRMEAS 8, 9, 12, 10
BHATT. 9, 8, 11, 12.

There was no real agreement; no one feature was in the top four of all three measures. This gave us confidence that there would be no overriding powerful feature to unbalance our selection. This is not the way to use the Bhattacharyya distance measure, however. The power of the Bhattacharyya measure is its ability to take features as a group. When we use the Bhattacharyya measure to select the best 4 of 12 features, the results change. Now the Bhattacharyya measure selects features 1, 6, 10, and 12 instead of 9, 8, 11, and 12. The questions to be answered now are, "How good are the selections?", and "How do they compare with OLPARS other measures?". For our comparisons we will compare the OLPARS measures of Fisher Pairwise Discriminant and Discriminant Measure against the Bhattacharyya measure. To check their relative performance we will use OLPARS's Decision Logic techniques of "Nearest Mean Vector", and "Fisher Pairwise" techniques in their Confusion Matrix, which gives the "percent correct" selections using the features selected by each measure. Table 5 gives the percent correct classifications, using the nearest mean vector (NMV), and Fisher pairwise (FPW) confusion matrix for the three Discriminant measures.

DISCRIMINANT MEAS.	FEATURES	NMV	FPW
Bhattacharyya	1, 6, 10, 11	88.0	97.6
Fisher	6, 10, 1, 2	82.8	93.9
Discriminant Meas.	8, 9, 12, 10	84.4	88.7

TABLE 5. BEST 4 OF 12 FEATURES

This test of the best four features selected by the three different techniques shows a dramatic result. The Bhattacharyya enhancement chose features different than the top four chosen by the other two OLPARS techniques, and its choices had a higher percentage of correct classifications according to both the NMV and the FPW logic!

We also decided to see how well it would do with poorer data. To do this we removed the best four features (1,6,10,and 11) from the set of 12 original features and would work with the remaining 8 features. In the next experiment we decided to select the best 3 features from 8 using the same three techniques. Table 6 presents the results of those measures.

DISCRIMINANT MEAS.	FEATURES	NMV	FPW
Bhattacharyya	5, 9, 12	84.9	91.7
Fisher	2, 4, 8	75.9	84.9
Discriminant Meas.	8, 9, 12	81.6	86.6

TABLE 6. BEST 3 OF 8 WORST FEATURES

Again the Bhattacharyya enhancement to OLPARS provides the best three features for producing the greatest number of correct classifications, by both evaluation measures! All similar trials on the NASADATA set provided the same result of superior performance when 2 or more features were considered together. This was a dramatic demonstration of the power of this new enhancement to OLPARS. The objectives of the Phase I activity had been met and exceeded. Not only would CNVEO have a technique that could chose the best pair of features, but a technique that would allow them to chose the best "n" of "L" features.

III. CONCLUSIONS

The KAB LABORATORIES INC. Team met or exceeded all Phase I objectives. The On-Line Pattern Recognition System (OLPARS) was delivered to CNVEO early, and quickly became operational on their computers. KAB provided an enhancement desired by CNVEO to select the best "pairs" of features. The KAB enhancement not only gives CNVEO the ability to select the best "pairs" of features, but more generally the best "n" out of "L" features. The KAB Team is also continuing work on two additional CNVEO desires, the identification of specific feature vectors, and the analysis of some CNVEO data. The original data did not come with sufficient information for analysis. These two additional products will be delivered shortly after the conclusion of the Phase I effort.

In the process of conducting the Phase I research, some observations were noted. CNVEO has a broad range of pattern analysis projects that can benefit from analysis aids. KAB has shown that it can develop and provide powerful aids to help this work. The OLPARS provides a powerful capability to perform feature set evaluation, but is not easily modified to perform additional analyses. A complementary system could further enhance CNVEO's capability to perform their work. That complementary system should complement OLPARS and provide a user-friendly, easy-to-modify, set of pattern analysis tools such as error measures, statistical measures, algorithm development aids and analysis aids.

The Phase I research also leads to a number of recommendations regarding a Phase II activity. Based upon KAB's successful research in Phase I, CNVEO now possesses a powerful feature set evaluation capability. That Phase I research demonstrated KAB's ability to provide CNVEO with useful analysis aids to support their work. To complete this process, a Phase II program should be initiated to permit KAB to develop the "missing pieces" required to round-out CNVEO's pattern analysis capabilities. A Phase II program would permit a complementary work-station to be developed which would assist CNVEO in conducting its other analysis efforts. This work-station, moreover, would find utility on a number of other U.S. Army and DoD research programs.